

GCACGTCGCA TGGAGACCA CCGTGAACGCCACCAAAATAT

 TGCCCAAGGTCTTACATAAGAGGACTCTTGGACTCTCAGC

 HNF4

 AATGTCAACGACCGACCTTGAGGCATACTTCAAAGACTGT

 HNF3-1 HNF

 TTGTTTAAAGACTGGGAGGAGTTGGGGAGGAGATTAGGT

 3-2

 TAAAGGCTTTGTACTAGGAGGCTGTAGG CATAAATTGGT

 CTGCGCACCAAGCACCATGCAACTTTTTCACCTCTGCCCTAA

 TCATCTCTTG
 Pre-genomic

* nucleotide conserved at >95% among 75 HBV strains

Fig. 1A

2701 TTATTATCCAGAACATCTAGGTAAATCATTACTTCCTAACTAGACCACTATTTACACACTCT
HNF1 HNF3

2761 ATGGAAAGGCGGGTTATATTTATATAAGAGAGAAACAACACATAGCGCCTCATTGTTGGGGTC
Sp1 TBP RNA Start

2821 ACCATATTCTTGGGAACAAGATCTACAGCATGGGGC
PreS1 protein start

Fig. 1B

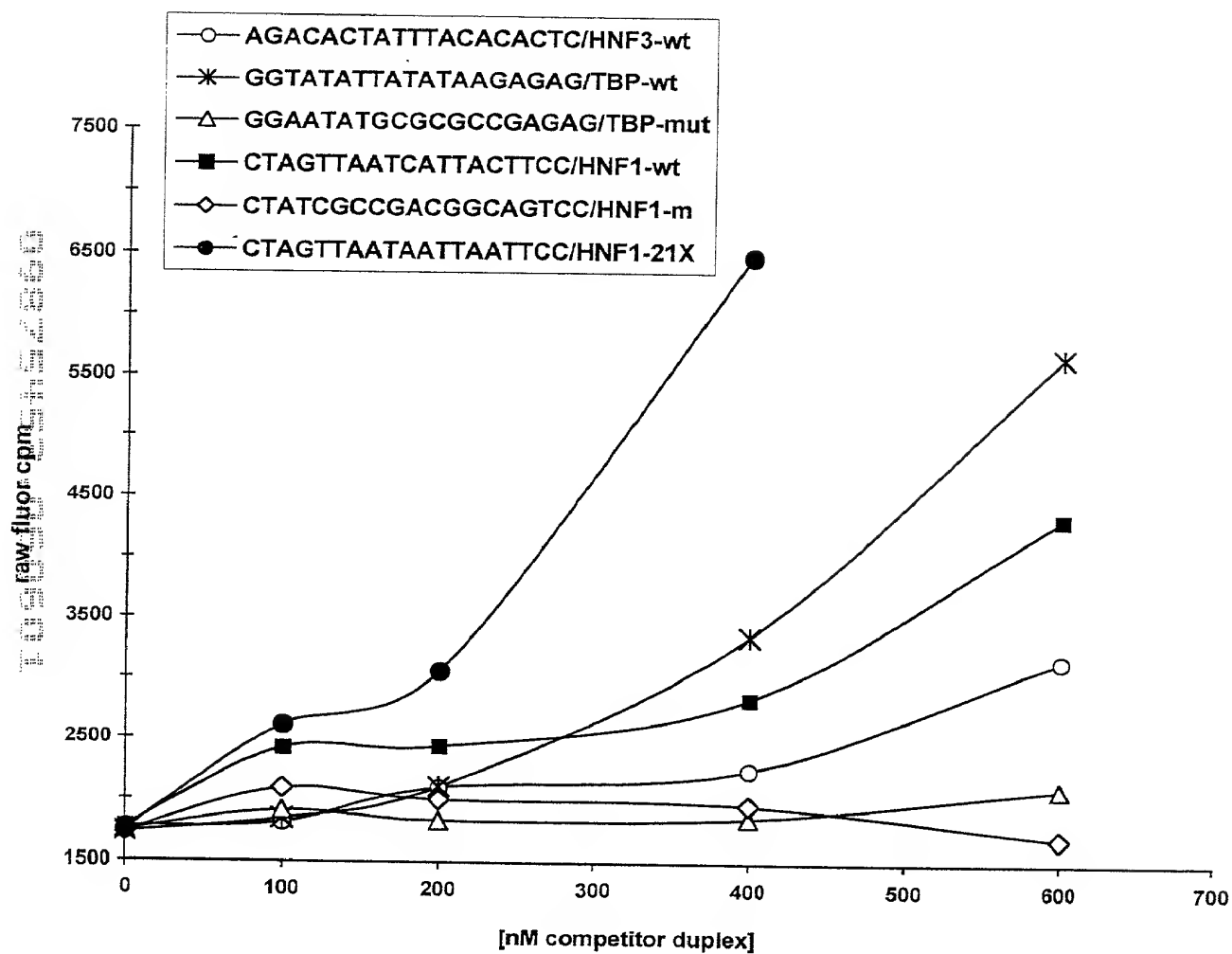


Fig. 2

	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030	2031	2032	2033	2034	2035	2036	2037	2038	2039	2040	2041	2042	2043	2044	2045	2046	2047	2048	2049	2050	2051	2052	2053	2054	2055	2056	2057	2058	2059	2060	2061	2062	2063	2064	2065	2066	2067	2068	2069	2070	2071	2072	2073	2074	2075	2076	2077	2078	2079	2080	2081	2082	2083	2084	2085	2086	2087	2088	2089	2090	2091	2092	2093	2094	2095	2096	2097	2098	2099	2100	2101	2102	2103	2104	2105	2106	2107	2108	2109	2110	2111	2112	2113	2114	2115	2116	2117	2118	2119	2120	2121	2122	2123	2124	2125	2126	2127	2128	2129	2130	2131	2132	2133	2134	2135	2136	2137	2138	2139	2140	2141	2142	2143	2144	2145	2146	2147	2148	2149	2150	2151	2152	2153	2154	2155	2156	2157	2158	2159	2160	2161	2162	2163	2164	2165	2166	2167	2168	2169	2170	2171	2172	2173	2174	2175	2176	2177	2178	2179	2180	2181	2182	2183	2184	2185	2186	2187	2188	2189	2190	2191	2192	2193	2194	2195	2196	2197	2198	2199	2200	2201	2202	2203	2204	2205	2206	2207	2208	2209	2210	2211	2212	2213	2214	2215	2216	2217	2218	2219	2220	2221	2222	2223	2224	2225	2226	2227	2228	2229	2230	2231	2232	2233	2234	2235	2236	2237	2238	2239	2240	2241	2242	2243	2244	2245	2246	2247	2248	2249	2250	2251	2252	2253	2254	2255	2256	2257	2258	2259	2260	2261	2262	2263	2264	2265	2266	2267	2268	2269	2270	2271	2272	2273	2274	2275	2276	2277	2278	2279	2280	2281	2282	2283	2284	2285	2286	2287	2288	2289	2290	2291	2292	2293	2294	2295	2296	2297	2298	2299	2300	2301	2302	2303	2304	2305	2306	2307	2308	2309	2310	2311	2312	2313	2314	2315	2316	2317	2318	2319	2320	2321	2322	2323	2324	2325	2326	2327	2328	2329	2330	2331	2332	2333	2334	2335	2336	2337	2338	2339	2340	2341	2342	2343	2344	2345	2346	2347	2348	2349	2350	2351	2352	2353	2354	2355	2356	2357	2358	2359	2360	2361	2362	2363	2364	2365	2366	2367	2368	2369	2370	2371	2372	2373	2374	2375	2376	2377	2378	2379	2380	2381	2382	2383	2384	2385	2386	2387	2388	2389	2390	2391	2392	2393	2394	2395	2396	2397	2398	2399	2400	2401	2402	2403	2404	2405	2406	2407	2408	2409	2410	2411	2412	2413	2414	2415	2416	2417	2418	2419	2420	2421	2422	2
--	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	---

Fig. 3

	CAGCTGGG	CCGCCCTTGT	GCGCGGGCTG	ATGCTCTGAG	GCTTGGCTAT
GCGGGGGCCA	ACGCGATTGT	GGGTGCTCGG	GGAGTGGGGG	GGGGCACGAC	CGTAGGTGCT
CCCTGCTGGG	GCAACCCATC	GCTCCCCATG	CGGAATCCGG	GGGTAATTAC	CCCCCAGGA
CCCGGAATAT	TAGTAATCCT	AATTCCCGGC	GGGGGAGGGG	GCGCGGGAGG	AATTACCCCT
GAAAGGTGGG	GGTGGGGGGG	GTCGCATCTT	GCTGTGAGCA	CCCTGGCGAA	GGGGAGAGGG
CTTTTTCTAT	CAGTTTTCTT	TGAGCTTTTA	CTGTTAAGAG	GGTACGGTGG	TTTGATGACA
CTGAACTATA	TTCAAAAGGA	AGTAAATGAA	CAGTTTTCTT	AATTTGGGGC	AGGTACTGTA
AAAATAAAAA	CAAAAGTTAA	GACAGTAAAA	TGTCCTTTTA	TTTTTTAATG	CACCAAAGAG
ACAGAACCTG	TAATTTTAAA	AACGTGTAT	TTAATTTAC	ATCTGCTTAA	GTTTGCGATA
ATATTGGGGA	CCCTCTCATG	TAACCACGAA	CACCTATCGA	TTTTGCTAAA	AATCAGATCA
GTACACTCGT	TTGTTTAATT	GATAATTGTT	CTGAATTATG	CCGGCTCCTG	CCAGCCCCCT
CACGCTCACG	AATTCAGTCC	CAGGGCAAAAT	TCTAAAGGTG	AAGGGACGTC	TACACCCCCA
ACAAAACCAA	TTAGGAACCT	CGGTGGTCTT	GTCCCAGGCA	GAGGGGACTA	ATATTTCCAG
CAATTTAATT	TCTTTTTTAA	TTAAAAAAA	TGAGTCAGAA	TGGAGATCAC	TGTTTCTCAG
CTTTCCATT	AGAGGTGTGT	TTCTCCCGGT	TAAATTGCCG	GCACGGGAAG	GGAGGGGGTG
CAGTTGGGGA	CCCCCGCAAG	GACCGACTGG	TCAAGGTAGG	AAGGCAGCCC	GAAGAGTCTC
CAGGCTAGAA	GGACAAGATG	AAGGAAATGC	TGGCCACCAT	CTTGGGCTGC	TGCTGGAATT
TTCGGGCATT	TATTTTATTT	TATTTTTTGA	GCGAGCGCAT	GCTAAGCTGA	AATCCCTTTA
ACTTTTAGGG	TTACCCCTT	GGGCATTTGC	AACGACGCCC	CTGTGCGCCG	GAATGAACT
TGCACAGGGG	TTGTGTGCCC	GGTCTCCCC	GTCTTGCAT	GCTAAATTAG	TTCTTGCAAT
TTACACGTGT	TAATGAAAAT	GAAAGAAGAT	GCAGTCGCTG	AGATTCTTTG	GCCGTCTGTC
CGCCCGTGGG	TGCCCTCGTG	GCGTTCCTTG	AAATGCGCCC	ATTCTGCCGG	CTTGATATG
GGGTGTGCGC	GCGCCCCAGT	CACCCCTTCT	CGTGGTCTCC	CCAGGCTGCG	TGCTGTGCCG
GCCTTCCTAG	TTGTCCCCTA	CTGCAGAGCC	ACCTCCACCT	CACCCCTTAA	ATCCCGGGGG
ACCCACTCGA	GGCGGACGGG	GCCCCCTGCA	CCCCTCTTCC	CTGGCGGGGA	GAAAGGCTGC
AGCGGGGCGA	TTTGCAATTC	TATGAAAACC	GGACTACAGG	GGCAACTCCG	CCGCAGGGCA
GGCGCGGCGC	CTCAGGGATG	GCTTTTGGGC	TCTGCCCCCTC	GCTGCTCCCG	GCGTTTGGCG
CCCGCGCCCC	CTCCCCCTGC	GCCCCCCCC	GCCCCCTCC	CGCTCCCATT	CTCTGCCGGG
CTTTGATCTT	TGCTTAACAA	CAGTAACGTC	ACACGGAATA	CAGGGGAGTT	TTGTTGAAGT
TGCAAAGTCC	TGGAGCCTCC	AGAGGGCTGT	CGGCGCAGTA	GCAGCGAGCA	GCAGAGTCCG
CACGCTCCGG	CGAGGGGCAG	AAGAGCGCGA	GGGAGCGCGG	GGCAGCAGAA	GCGAGAGCCG
AGCGCGGACC	CAGCCAGGAC	CCACAGCCCT	CCCCAGCTGC	CCAGGAAGAG	CCCCA

Fig. 4

10	20	30	40	50	60	70
GAATTCAC	TGGGAGAG	TCAGGAAG	GACAACAG	TAATAGGT	ACAGAGTA	ATAGAGGT
CTTAAGTG	ACCTCTCG	TAGTCTTCT	CTGTTGTC	CTTATCCAG	TGTCTCAT	TCTCTCCAG
80	90	100	110	120	130	140
CTAAAAATA	ACTCTAAG	GTATTCAG	AAAACATA	TTGAGCTA	AATGGTGG	TCAATTTCA
GATTTTTAT	TGAGATTCT	CATAAGTC	TTTTGATA	AACTCGAT	TTACCACCT	AGTTAAAGT
150	160	170	180	190	200	210
GGGAATATT	TGGGCAGA	TCAGACTG	GGAGGCTG	GATCAAGA	TTGAGGCA	GAGGTTGG
CCCTTATA	ACCCGTCT	AGTCTGAC	CCTCCGACC	CTAGTTCT	AACTCCGT	CTCCAACCT
220	230	240	250	260	270	280
AACAACGT	TTTTCAAG	GGTCACGT	ACAAATCT	GACCTTCAG	CTCCCCCT	TCGGGTCT
TTGTTGACA	AAAAGTTCA	CCAGTGCA	TGTTTAGAC	CTGGAAGTC	GAGGGGAG	AGCCAGAAC
290	300	310	320	330	340	350
GCTGAGCT	TTGCAGGG	CCTGCAGCT	TGGCACTCT	AAGTTGTAT	AACTGACAG	TGCAGAAGT
CGACTCGAC	AACGTCCCG	GGACGTGAG	ACCGTGAG	TTCAACAT	TTTGACTGT	ACGTCTTCA
360	370	380	390	400	410	420
CTTGAGCCC	TTTTGGCT	CATGATAAT	TTCCTTCAG	GGAACATA	TTACTTGT	AAGAACCAG
GAACTCGGG	AAAACCGAG	GTACTATT	AAGGAAGTC	CCTTGATT	AATGAACAG	TTCTTGGTT
430	440	450	460	470	480	490
GCCTCTGAC	TGACTGAT	AAGTTCAT	CGTGCATCG	AGCCACCT	TTGGCAGAT	TAGTGAAA
CGGAGACTG	ACTGACTAG	TTCAAGTAG	GCACGTAG	TCGGTGAT	AACCGTCT	ATCACTTTT
500	510	520	530	540	550	560
CTACATAG	CTGGGCCC	GACAGGAT	TGGGGCGTG	GAGGGGAA	AAGCAGGT	TAAGTATAT
GATGTATCT	GACCCGGGT	CTGTCTTAC	ACCCCGCAC	CTCCCCCT	TTCTGTCC	ATTGATAT
570	580	590	600	610	620	630
GATAGCAT	CTATCAGAG	AGTTTTTAC	TTTCTTATT	GTCTCTCAA	ACAATTTT	AGGAATCAT
CTATCGTAC	GATAGTCT	TCAAAAAT	AAAGGATA	CAGAGAGTT	TGTTAAAAT	TCCTTAGTA
640	650	660	670	680	690	700
AAAGCAATT	TATCATGG	TCTAGACC	GTTTGGAT	GAGGTAGGG	TTTCCACAG	TGCTTTTAG
TTTCGTAAA	ATAGTACCA	AGATCTGG	CAAACCTAC	CTCCATCC	AAAGGTGTC	ACGAAAAT
710	720	730	740	750	760	770
TTGAAGGAA	TCTGATAAG	TGATGCAAAA	GCCCTTCAG	AATGTGTA	CCTACACAC	TCAGTGATT
AACTTCCTT	AGACTATT	ACTACGTTT	CGGGAAGTC	TTACACATT	GGATGTGT	AGTCACTAA
780	790	800	810	820	830	840
AATTCATT	CAAAACTTA	GGTGTTTT	ATATTGTT	TGTTCAATT	GTTTTTACC	ACATGTAAG
TTAAGTAAC	GTTTTGAAT	CCACAAAA	TATAACAAT	ACAAGTAA	CAAAAATGG	TGTACATT
850	860	870	880	890	900	910
AGTTGGCA	TATTTGTAA	ACTCATGT	TAGGCTAA	AAATTCCAA	AAATTCAG	TGAGAATT
TCAACCGTT	ATAACAATT	TGAGTACAG	ATCCGATTT	TTTAAGGTT	TTTAAGTC	ACTCTTAAC

Fig. 5A

920	930	940	950	960	970	980
TTATTGCTTA	ACGTGTGTCA	AATTTCTTCC	ATGCACATCT	TTATTAGATC	TTCACAGCAA	CCTACAGGAT
AATAACGAAT	TGCACACAGT	TTAAAGAAGG	TACGTGTAGA	AATAATCTAG	AAGTGTGCTT	GGATGTCCTA
990	1000	1010	1020	1030	1040	1050
AAGCAAGACA	GGTGCAAGTG	CCTCCTTTGG	GTATGAGGAA	ACTGAGGTCT	AAAGAGATGA	AGTGATTTGC
TTCGTTCTGT	CCACGTTTAC	GGAGGAAACC	CATACTCCTT	TGACTCCAGA	TTTCTCTACT	TCACTAAACG
1060	1070	1080	1090	1100	1110	1120
CCAAGGCTCA	TAGCAATTTA	TTGGTAGAGC	AAAGACTAGA	ATTCTCTTAA	CTGCAGCCTA	TTTTCCCTAT
GGTTCGAGT	ATCGTTAAAT	AACCATCTCG	TTTCTGATCT	TAAGAGAATT	GACGTCGGAT	AAAAGGGATA
1130	1140	1150	1160	1170	1180	1190
TCTGAACGTG	TACATCAGCA	TCAACAATTA	TCTAATGGAT	TGGAACAGTG	TACACAGGCA	GCTTAGCTAC
AGACTTGACA	ATGTAGTCGT	AGTTGTTAAT	AGATTACCTA	ACCTTGTCAC	ATGTGTCCGT	CGAATCGATG
1200	1210	1220	1230	1240	1250	1260
GTCAAGTCAC	GATTTTACT	TTAACTTCAA	TTCCAGAGTC	TTGGCCTGAT	TTCCCTCAAG	ACCCTACTTA
CAGTTCAGTG	CTAAAAATGA	AATTGAAGTT	AAGGTCTCAG	AACCGGACTA	AAGGGAGTTC	TGGGATGAAT
1270	1280	1290	1300	1310	1320	1330
TCTTTGGCTT	TGGAAAAATT	ATTTTCTTG	CATTATCTTT	CCAGCTAAAT	TTTATTTAAT	AACCATCAGC
AGAAACCGAA	ACCTTTTAAA	TAAAAAGAAC	GTAATAGAAA	GGTCGATTTA	AAATAAATTA	TTGGTAGTCG
1340	1350	1360	1370	1380	1390	1400
ATGCTTTTTT	TGCTTTATGC	CATGTAGACT	TGACCTGAAA	ACCTGCCAGG	CTTTCATTGA	GTTTAGTGAT
TACGAAAAAA	ACGAAATACG	GTACATCTGA	ACTGGACTTT	TGGACGGTCC	GAAAGTAACT	CAAATCACTA
1410	1420	1430	1440	1450	1460	1470
TAAAGAAGTA	AAGTTCGAG	AAGCAATTAG	TTGATGGGAC	ACCAGTCATA	AAATCAATCC	AAACTTTTGT
ATTTCTTCAT	TTCAAGACTC	TTCGTTAATC	AACTACCCTG	TGGTCAGTAT	TTTAGTTAGG	TTTGAAAACA
1480	1490	1500	1510	1520	1530	1540
TGACATGTGT	TTCTTTCTCC	ATATACCAGG	TTCCCGCTTC	GTATTAGTAA	GATTGAAATT	GAAATAAGTC
ACTGTACACA	AAGAAAGAGG	TATATGGTCC	AAGGGCGAAG	CATAATCATT	CTAACTTTAA	CTTTATTGAG
1550	1560	1570	1580	1590	1600	1610
TATTGCTGGT	GGATGAATTT	GTCACTTTCC	TTGAAACTGG	TGAACCCAAA	AAGTTAGACA	GTGATAGGAA
ATAACGACCA	CCTACTTAAA	CAGTGAAAGG	AACTTTGACC	ACTTGGGTTT	TTCAATCTGT	CACTATCCTT
1620	1630	1640	1650	1660	1670	1680
AATACTGCCA	TTGTCTGTTA	AGAAGTCTAT	GACATTTCAA	GGCAAGAATG	AATATATGGA	AGAAGAACT
TTATGACGGT	AACAGACAAT	TCTTCAGATA	CTGTAAAGTT	CCGTTCTTAC	TTATATACCT	TCTTCTTTGA
1690	1700	1710	1720	1730	1740	1750
TGTTTCTTCT	TTACTTACAA	AAAGGAAAGC	CTGGAAGTGA	ATGATATGGG	TATAATTAAA	AAAAAAAAAA
ACAAAGAAGA	AATGAATGTT	TTTCCTTTTCG	GACCTTCACT	TACTATACCC	ATATTAATTT	TTTTTTTTTT
1760	1770	1780	1790	1800	1810	1820
AAAACAAAAA	ACCTTTACGT	AACGTTTTGC	TGGGAGAGAA	GACTACGAAG	CACATTTTCC	AGGAAGTGTG
TTTTGTTTTT	TGGAAATGCA	TTGCAAAACG	ACCCTCTCTT	CTGATGCTTC	GTGTAAAAGG	TCCTTCACAC

Fig. 5B

1830	1840	1850	1860	1870	1880	1890
GGCTGCAACG	ATTGTGCGCT	CTTAACTAAT	CCTGAGTAAG	GTGGCCACTT	TGACAGTCTT	CTCATGCTGC
CCGACGTTGC	TAACACGCGA	GAATTGATTA	GGACTCATTG	CACCGGTGAA	ACTGTCAGAA	GAGTACGACG
1900	1910	1920	1930	1940	1950	1960
CTCTGCCACC	TTCTCTGCCA	GAAGATACCA	TTTCAACTTT	AACACAGCAT	GATCGAAACA	TACAACCAAA
GAGACGGTGG	AAGAGACGGT	CTTCTATGGT	AAAGTTGAAA	TTGTGTCGTA	CTAGCTTTGT	ATGTTGGTTT
1970	1980	1990	2000	2010	2020	2030
CTTCTCCCCG	ATCTGCGGCC	ACTGGACTGC	CCATCAGCAT	GAAAATTTTT	ATGTATTTAC	TTACTGTTTT
GAAGAGGGGC	TAGACGCCGG	TGACCTGACG	GGTAGTCGTA	CTTTTAAAAA	TACATAAATG	AATGACAAAA
2040	2050	2060	2070	2080	2090	2100
TCTTATCACC	CAGATGATTG	GGTCAGCACT	TTTTGCTGTG	TATCTTCATA	GAAGGCTGGA	CAAGGTAAGA
AGAATAGTGG	GTCTACTAAC	CCAGTCGTGA	AAAACGACAC	ATAGAAGTAT	CTTCCGACCT	GTTCCATTCT
2110	2120	2130	2140	2150	2160	2170
TGAACCACAA	GCCTTTATTA	ACTAAATTTG	GGGTCCTTAC	TAATTCATAG	GTTGGTTCTA	CCCAAATGAT
ACTTGGTGTT	CGGAAATAAT	TGATTTAAAC	CCCAGGAATG	ATTAAGTATC	CAACCAAGAT	GGGTTTACTA
2180	2190	2200	2210	2220	2230	2240
GGATGATGGT	AGAAACCAAA	TAGAAGAATG	GTCTTGTTGG	ATAATGTTTG	TTCCCTAGTC	AATGAACTCT
CCTACTACCA	TCTTTGGTTT	ATCTTCTTAC	CAGAACACCG	TATTACAAAC	AAGGGATCAG	TTACTTGAGA
2250	2260	2270	2280	2290	2300	2310
CATATTCTTG	TCTCTGGTTA	GGATCTTGGG	ATCTGGAGTC	AGACTGCCTG	GGCTCAAATC	TTGGCTCTGC
GTATAAGAAC	AGAGACCAAT	CCTAGAACCC	TAGACCTCAG	TCTGACGGAC	CCGAGTTTAG	AACCGAGACG
2320	2330	2340	2350	2360	2370	2380
CCATACCATC	TCTGTTATCC	TGGGGCAAGT	GCCTCAGTTT	CCACATCTGA	GAAATGGGGA	TGGTAGTGGT
GGTATGGTAG	AGACAATAGG	ACCCCGTTCA	CGGAGTCAAA	GGTGTAGACT	CTTTACCCCT	ACCATCACCA
2390						
GTCCATTTCA	TAGAT					
CAGGTAAAGT	ATCTA					

Fig. 5C

GAGATGTATATAATTTTTTAGGAAAATCTCAAGGTTATCTTTACTTTTTCTTA
GGAAATTAACAATTTAATATTAAGAAACGGCTCGTTCTTACACGGTAGACTTA
ATACCGTAAGAACGAGCCGTTTTTCGTTCTTCAGAGAAAGATTTGACAAGATTA
CCATTGGCATCCCCGTTTTATTTGGTGCCTTTCACAGAAAGGGTTGGTCTTAA
TT

Fig. 6

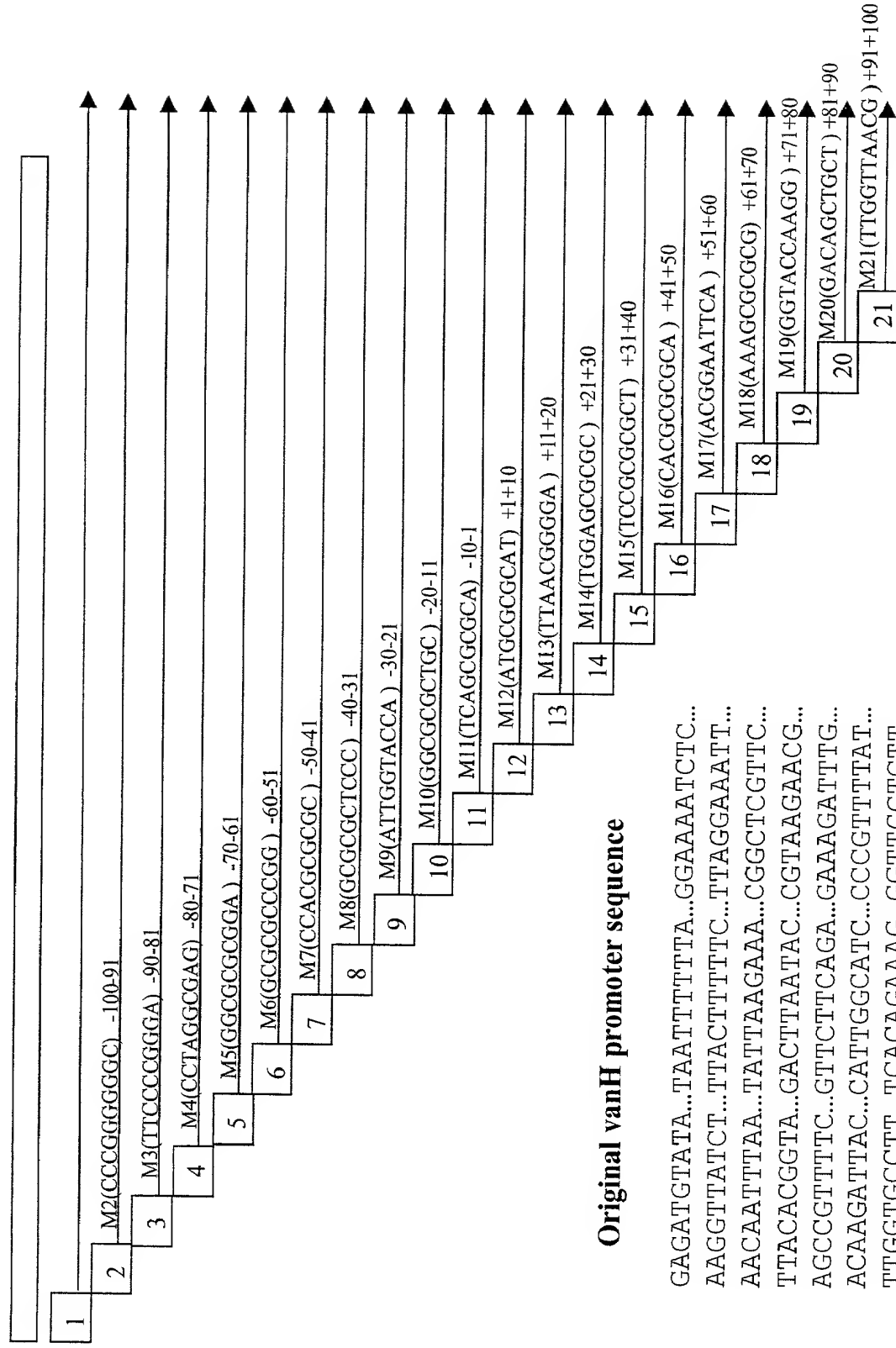


Fig. 7

TCTAGAAAAT	AATTCCCAAT	ATTGAATCCC	AAAGAATTCA	ACATTTGGGC	TGTCGTTTGA	61
AAGATAAGTT	GAATTTGGTC	ATGAAGGAAG	AGAGGGGGGA	TACAAATTCA	GTAAGAGGTA	121
ACAGCAAGGT	CCAAAGACAG	TCAGGTCTTC	AGTAGTATGG	AGTATATTCA	GAGGGAGCCA	181
AGATGTCTGA	TGTGAACTAA	AAAGATTGGT	GGTTGGTAGG	AGGAAGAGGT	GTGAGAAGAG	241
GCTGTAAAGA	AAAATTGAAA	CTTGATTGTG	ATGGACTTTA	AAGGCTAGGC	TATGGGACTT	301
GGACATGAAT	CTGCAGGCCA	GTGTTTGCAG	ACTGGCGCCC	ATAACTGTCT	ATCACAGCAA	361
CACAGACATG	TGTTGTTTGG	CCTGCAGAGG	TTTGGCCTGC	ATGATGATTT	TAAACCATCT	421
GAATTAGTAG	CCATCATTTT	CAAAAATCAA	GAGATGCCAC	ATTAAAAATAT	GGAATGCTGC	481
TGTTCTTGAA	AATAATGAAA	CATCTGGAAC	ATTGAGGCCA	CATTCCTGAC	TGACAGCAAT	541
CAGTTGGAGC	TGCGTAGTGA	CTGCCCCACTT	TACATGGGGC	ATCTGATCCC	TAGTCGATTA	601
CAGCTGCCAC	CACTTCCCTT	TATCTCTCTA	ATACCAAGCT	CTTTTCACTC	ATTTTTGTTA	661
CTTAAGAGAT	ATTTGGGTTT	GAAACCTCTG	ATGCAGGTAA	TTGAGGGTTA	TAGAGCAGAG	721
GACAGATGCT	ATCAGAGTTG	TCTTTTAAGA	AAGAACCCTC	TGTTCTTCAT	TTTGTGGAAG	781
ATAGCCTGGA	AGAGGGCAGC	CAGGGGAGAA	GTTAGGGCTG	GAGCTATGAG	AAAGCATAAG	841
ATGAGATGAT	GGCTTCAACA	TTGAGGACAG	AAAGAATATT	GAGATGAGAA	AGTAGTCCAT	901
ATAAGCATCT	ATGCAAAGGA	AATAGCAGAT	GTCTTCAAAT	CAGCAGAGGC	AACAACCTCTG	961
AAAGTTTATT	CATAAGCCCC	TCTTTTCATC	TCCAATCCAG	TTCAAATGTA	ATTATTTAAA	1021
TTGTTCTTCA	CTCTCCTTCC	TGGATCATGA	ATGAGCTCCT	TAAATGCAGG	GTCCACAGTG	1081
TCCTATTTCAT	CAGTGAATTC	CAAGTGCCTA	GCACAGAGCC	TGGCAAATAG	TAAATGCTTA	1141
ACAAATATTC	GTTCAGTGCA	TGAATTGGAG	TGATTCTCTA	CTTTGCCTCA	TAAGTTGAAA	1201
AAAGGTTTAT	TACATACCTA	AATATGCTGA	AATCACAGGG	CATTTGGCAA	CCCCCCCCAA	1261
CCAAAACCTCC	CAGTTTGGAA	ACAGAATTTT	AATTCTGTGA	AAATAAAATC	CATTCATTTA	1321
TTCAAAAAAT	ATTTATTAAA	CAATGACCAT	GTCCACACCA	GGCTGAGTCC	TAAGGATTCA	1381
ATGATGAACA	AAAACCAACA	TGATTCTCTG	TCTTAGGAAA	CATACAGTTC	AGTGAGGAAA	1441
ACAGATTGTG	AGAAGTCCTC	CAACAAATAC	TGGGTGCTAT	TAAAATATAT	TAAAAGGTGA	1501
GTGGGTGAGG	GACTTGAGCT	AGCCTAGGTG	GTTCAGGAAG	TCTTCCTGGA	TGTGCTGATA	1561
TGCATAGGCA	TTAACTAGAT	AAATAGAGAG	AAGGATGAAC	CAACATTGCA	GGTAGAGGGA	1621
ACAGAATATG	CAAAGGCAGG	AAGGATTATG	GAGTCGTTGG	AGGACCTGAA	TAAAGGCCCA	1681
GTGTAAGTGG	ATCTCAGAAA	ACAGGAGGAA	AGGTGTATGA	GATGAGATCA	GAGAGGCAGA	1741
TCATGTGGGG	TATGGTTAAT	GTTTTGGACT	TTTCTATTAA	GAGCAATGGG	GAGACAGTGA	1801
CAGGACTTAA	ACGGGGAAAT	AATATGACCA	GATTAAACTT	TCTAAAAAAC	CCTCTATGCA	1861
AATATATATT	GAGAGTTAAT	TATTGACAAA	GATTCAAAGG	CAACAAAGTG	GAGAGAGAAT	1921
AGTATTTTCA	AAAAATGGTG	CCAAAACAAT	AGGACATCTA	TATTAAAAAGT	TGGGTATCTG	1981
TCTACAAAAC	TTAATTCAAA	ATGGATCACA	GACCTAAATG	TAAAACCTGAA	AGCTATACAA	2041
CTTCTGGAAG	GAAAACACAG	ATGGGAATCT	GTGTGATCTT	GAGTTTGAAA	ATGATTTTATT	2101
ATATCTGACA	CCATAATCCG	TAAGTTAACA	TAATTCATAA	GTGAACAAAG	TGATGAACTG	2161
GACTTCATCA	GAATTTAAAA	TGTTTGTGCT	TCAAAAAGACA	CTGGTATGAT	AATGAAGACA	2221
AATACAGAT	AAGATATTGT	TGAATCATAT	TTCTGATAAA	GGAATTGTGG	CTCAGAATAC	2281
ATAACTCTAA	ACCCCATAA	TAAATTACAA	GTAGCCCAAT	TAAAAAATAA	AAAAGAGAAA	2341
AAATTTACAG	TCTTCATCAA	AGAAAGTATC	AATTGTAAAA	TAAGCACATG	AAAAATGCTC	2401
TGCATCTTTA	TTCATGGGGG	GATGAAATAA	AAATTAAATG	GGAAAGACAC	CTCTAATTAG	2461
AATACTAAAA	TTAAAAAGAC	TGACCATACC	AAGTATTGGT	GAAGTGGAAG	TGTAAAATGA	2521
TACAATCAAC	TTAGGTAGAT	GATTTGGAAG	TTTCTTACAA	AAGTAGGTGT	ATACCTACCC	2581
TGTGACTCAC	CCATTCCATG	GCTAAGTATT	TACCTGAGAG	AAATGAAAGA	ATACATCCAT	2641
ACAAAGATGT	TTATACAAAT	ATTTATAGCA	GTTTTATTTG	TAGTAGCCCC	AAACTGAAAA	2701
GAACCCAAAT	GTCCATCAAA	AGTGAATGGA	TAAACAAAGC	GTGGTACAGC	AATGCAATAG	2761
AATACTACTT	AGCAATAAAG	AAGAATGAGC	TAGTGATATA	CATAACAGCT	TAAATGTACA	2821
TCAAAGGCAT	TGTGCTCAGT	GAAAGATGCA	AGTAAAAAAA	AAAAAGAGTA	CATGCTGTAT	2881
AGTTCCATTG	ACATAAAACT	CTGGAAAAGT	AAAAACAGTC	TATACTGACA	GAAAGCAGAT	2941
CATTGGTTGC	CTGAGGAGGA	GGAGTATAGG	AGAGGTGGAG	GGAAAATGTA	CAAAGTGGCA	3001
CAATAAAAAAC	TTTTGGAATC	ATAGATATAT	TCACTATCTT	GATTGAGTGA	TGATTTTCATG	3061

Fig. 8A

AGTGCACGTG	CGTGTGTCAA	AAATGATCAA	TTTATGCAAC	TTTAAATATG	TGCAGTTTAT	3121
TGTATATATC	AATTATACCT	CAGTACGGCT	ATTAAAAAGA	AACCCCTCTGG	CTGCACAATG	3181
CAGAACTGAT	TCTAGGAAAG	AGTGGAGGGA	GGATGACCAT	TTACAGTGCT	CCAGGTGGAA	3241
GAGAACGGTG	CCTTCTGGAA	GTGAACTAGG	TTGGCAACAA	CAGAGATGAA	ATAAATGGGC	3301
AGATGTGTGA	GATACTTAGG	AAATAAAACC	CGATGGTCAC	CATTTTCCAA	AGGTCAGCTC	3361
ATCCTGGCTT	TCCAGAGCAA	AGAGCTAGGG	AAGACTTTAT	TAATAAATCC	CTCTTGAAGT	3421
TGCAGAGGAA	GCTTATAGCA	GAAACTTACT	CTCAACCTGA	CTAATCTGAG	AGAACACCTC	3481
TGGTTCCATT	TGATTACTAA	AAAACCTGCA	AGAACAGGAG	GAGAAAGAAG	AAGAAAGCTG	3541
GTACAAACAG	TGAACTTATA	TAATATTAAT	CAATAATTGT	CTCTTGTCT	TAAAAGCAAT	3601
GGGAAGAAAA	TGAGATTTGA	GCTGGAAGAT	CAGAGTTCAA	AATCCAAATA	AAGTATATGG	3661
CCCTAATATG	CTTATAGTAG	TTAACCTTTC	CTGATAATGA	TATAATTGTT	GACAGCACCA	3721
TCTTTAAAAA	AAAATAACAT	AGTAATCCTT	CAGATTTGTA	GAAGATCTTT	CCTGTTTACA	3781
AGTTTGTCTT	ATACACATTA	TGCTTTTAA	ATGACACACT	AGCCTTCTGA	GGGTAACTTA	3841
TATTGGCAAC	AGTTTTTCAGA	TGTGGAAACT	GTGAAGACAA	TGTTGGTGAT	GTGGAAGCAA	3901
CATAAACTTT	GGAGTCTTTC	AGACCCAGGT	TTGAATGTCA	GACTGCTTTT	TATTCAGAGT	3961
AACTTCAGAG	CATTATTTCT	CACCTTAATT	TTTTTTCAGG	CCTCTTTGTG	TCTATGTGTC	4021
CTCTTCACTC	CTGTCCATTG	TTTCTTCAGT	GATTTTTGCC	ACCTTCCTTC	ACTGTTAGTG	4081
TGTAGACACA	TAGTTCCTCT	GGCTCTGAGA	GCCTATGTTA	ATTCCATTCT	ACCATCCTGC	4141
CACGGCCAC	TCAATTCCCTA	TTGAGCAATG	CTAGTTGAAA	GTTGTGGTGG	GATTAAATGT	4201
TGCAATGAGT	ATTCAAATGA	GGTTGAAGTA	TCTACGCATT	CTACTTACAT	ATGGTGAGGT	4261
ATATTCAAGG	AAGCTGTAGC	CATTAANAATC	TCAGGAAATA	ATTTTTTACC	TCCTCAGGTG	4321
AAAGGGTCTT	CAGGCCTTTG	TGTTCTGGAA	GGTTCATTTA	TAGCCATTTT	CCAAATGACA	4381
ATGCGATTGA	TGAGTCTAGA	GTCTAGCTCA	AATAGCAATG	GACTGGAAGA	CTAGTTTAGG	4441
TTTTACTAAT	GTGGAACATA	GAACAAATTA	TGTCCTTGTT	TCAGCCTGTT	CATCTGTGAA	4501
ATAGAGCCTA	TCATATCCAG	TCTTCCTTGC	CTTTAGGTTT	GAGTTACCTT	CTTTGGTCAA	4561
GGTAAGTAAA	TGCCTATGAT	GTTTGGCTGT	GCACAAGATA	AAGCTACAAC	AAAGCTACAA	4621
CCCATCTTTT	CTCTGTAGAA	GACTCAAAAA	GCAAAAGAGA	CCCAGGAAAA	TCTCGGAATG	4681
ACTTTTGGAA	CAGAGAGCCT	CCCCAGAATC	AGAAGTCAAG	GAATTTAAAC	ATAGGGAAGG	4741
CCCAGGTCTC	TACTGACATA	AAGGAAAGAT	GTTTTCTTAT	AGGTTTCACG	TTTACATTTT	4801
CTCTCTCTTG	ATCCCATTCC	CACCTGCATC	TGCCACCTTT	ACACAGGGCT	TATGGGACCT	4861
CCTCCACAAA	AGAGCAGTTG	CAGTAACCCA	CATCATCCTC	TACGCCCTGG	CTGTCCATCA	4921
AGAGGCGAAA	AGCAGCCCTA	TATAGGTTCT	ATCCTTGGAT	AGTTCCAGTT	GTAAAGTTTA	4981
AAATATGCGA	AGGCAACTTG	GAAAAGCAAG	CGGCTGCATA	CAAAAGCAAAC	GTTTACAGAG	5041
CTCTGGACAA	AATTGAGCGC	CTATGTGTAC	ATGGCAAGTG	TTTTTAGTGT	TTGTGTGTTT	5101
ACCTGCTTGT	CTGGGTGATT	TTGCCTTTGA	GAGTCTGGAG	AGTAGAAGTA	CTGGTTAAAG	5161
GAACCTCCAG	ACAGGAAGAA	GGCAGAGAAG	AGGGTAGAAA	TGACTCTGAT	TCTTGGGGCT	5221
GAGGGTTCCCT	AGAGCAAATG	GCACAATGCC	ACGAGGCCCG	ATCTATCCCT	ATGACGGAAT	5281
CTAAGGTTTC	AGCAAGTATC	TGCTGGCTTG	GTCAATGGCTT	GCTCCTCAGT	TTGTAGGAGA	5341
CTCTCCCACT	CTCCCATCTG	CGCGCTCTTA	TCAGTCCTGA	AAAGAACCCC	TGGCAGCCAG	5401
GAGCAGGTAT	TCCTATCGTC	CTTTTCCTCC	CTGCCCTCCA	CCACCCTGTT	GGTTTTTTAG	5461
ATTGGGCTTT	GGAACCAAAT	TTCCTGAGTG	CTGCCCTCCA	GGAAATCTGG	AGCCCTGGCG	5521
CCTAAACCTT	GGTTTAGGAA	ACCAGGAGCT	ATTGAGGAAG	CAGGGGTCCT	CCAGGGCTAG	5581
AGCTAGCCTC	TCCTGCCCTC	GCCCACGCTG	CGCCAGCACT	TGTTTCTCCA	AAGCCACTAG	5641
GCAGGCGTTA	GCGCGCGGTG	AGGGGAGGGG	AGAAAAGGAA	AGGGGAGGGG	AGGGAAAAGG	5701
AGGTGGGAAG	GCAAGGAGGC	CGGCCCGGTG	GGGGCGGGAC	CCGACTCGCA	AACTGTTGCA	5761
TTTGCTCTCC	ACCTCCCAGC	GCCCCCTCCG	AGATCCCAGG	GAGCCAGCTT	GCTGGGAGAG	5821
CGGGACGGTG	CGGAGCAAGC	CCACAGGCAG	AGGAGGCGAC	AGAGGGAAAA	AGGGCCGAGC	5881
TAGCCGCTCC	AGTGCTGTAC	AGGAGCCGAA	GGGACGCACC	ACGCCAGCCC	CAGCCCGGCT	5941
CCAGCGACAG	CCAACGCCTC	TTGCAGCGCG	GCGGCTTCGA	AGCCGCCGCC	CGGAGCTGCC	6001
CTTTCCTCTT	CGGTGAAGTT	TTTAAAAGCT	GCTAAAGACT	CGGAGGAAGC	AAGGAAAGTG	6061

Fig. 8B

CCTGGTAGGA	CTGACGGCTG	CCTTTGTCCT	CCTCCTCTCC	ACCCCGCCTC	CCCCACCCT	6121
GCCTTCCCCC	CCTCCCCCGT	CTTCTCTCCC	GCAGCTGCCT	CAGTCGGCTA	CTCTCAGCCA	6181
ACCCCCCTCA	CCACCCTTCT	CCCCACCCGC	CCCCCCGCCC	CCGTCGCCCA	GCGCTGCCAG	6241
CCCGAGTTTG	CAGAGAGGTA	ACTCCCTTTG	GCTGCGAGCG	GGCGAGCTAG	CTGCACATTG	6301
CAAAGAAGGC	TCTTAGGAGC	CAGGCGACTG	GGGAGCGGCT	TCAGCACTGC	AGCCACGACC	6361
CGCCTGGTTA	GGCTGCACGC	GGAGAGAACC	CTCTGTTTTC	CCCCACTCTC	TCTCCACCTC	6421
CTCCTGCCTT	CCCCACCCCG	AGTGCGGAGC	CAGAGATCAA	AAGATGAAAA	GGCAGTCAGG	6481
TCTTCAGTAG	CCAAAAAACA	AAACAAACAA	AAACAAAAAA	CAAGAAATAA	AAGAAAAAGA	6541
TAATAACTCA	GTTCTTATTT	GCACCTACTT	CAGTGGACAC	TGAATTTGGA	AGGTGGAGGA	6601
TTTTGTTTTT	TTCTTTTAAG	ATCTGGGCAT	CTTTTGAATC	TACCCTTCAA	GTATTAAGAG	6661
ACAGACTGTG	AGCCTAGCAG	GGCAGATCTT	GTCCACCGTG	TGTCTTCTTC	TGCACGAGAC	6721
TTTGAGGCTG	TCAGAGCGCT	TTTTGCGTGG	TTGCTCCCGC	AAGTTTCCTT	CTCTGGAGCT	6781
TCCCGCAGGT	GGGCAGCTAG	CTGCAGCGAC	TACCGCATCA	TCACAGCCTG	TTGAACTCTT	6841
CTGAGCAAGA	GAAGGGGAGG	CGGGGTAAGG	GAAGTAGGTG	GAAGATTCAG	CCAAGCTCAA	6901
GGATG						

Fig. 8C

